Class 4: Dual problems and weak duality theorem

In *duality theory*, we associate an optimization problem with the so-called dual problem that, under some conditions, can be equivalent in some sense to the original problem. This equivalence can be established in linear programming problems, but even for general optimization problems, the duality theory is known to be useful. As usual, we consider the following nonlinear programming problem:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X, \end{array} \tag{NLP}$$

where

$$X := \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}. \tag{1}$$

Once again, we assume that $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g\colon \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable, and we write $g := (g_1, \dots, g_m)$ with $g_i\colon \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$. In this context, we call (NLP) the *primal problem*.

Consider the following *Lagrangian function* $L\colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ associated to (NLP):

$$\begin{aligned} L(x, \lambda, \mu) &:= f(x) + \langle g(x), \lambda \rangle + \langle h(x), \mu \rangle \\ &= f(x) + \sum_{i=1}^{m} g_i(x)\lambda_i + \sum_{j=1}^{p} h_j(x)\mu_j, \end{aligned}$$

where $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ are called Lagrange multipliers associated to the inequality and the equality constraints, respectively. We first show the following result, that will be used to reformulate the primal problem.

Proposition 1. *For problem (NLP), the following equality holds:*

$$\sup_{(\lambda,\mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p} L(x, \lambda, \mu) = \begin{cases} f(x), & \text{if } x \in X, \\ \infty, & \text{otherwise.} \end{cases} \tag{2}$$

Proof. We prove separately for the cases that (i) x is feasible and (ii) x is not feasible.

(i) Assume that $x \in X$. For all $(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$, since $\langle g(x), \lambda \rangle \leq 0$ and $\langle h(x), \mu \rangle = 0$, we obtain

$$L(x, \lambda, \mu) \leq f(x).$$

Moreover, we have $L(x, 0, \mu) = f(x)$ for all $\mu \in \mathbb{R}^p$. Therefore, in this case, (2) holds.

(ii) Assume that $x \notin X$. We also have two cases to consider: (a) there exists $j \in \{1, \dots, p\}$ such that $h_j(x) \neq 0$, or (b) there exists $i \in \{1, \dots, m\}$ such that $g_i(x) > 0$.

(a) Assume that $h_j(x) \neq 0$ for some $j \in \{1, \dots, p\}$. Then, we can define

$$\mu_i = \begin{cases} th_i(x), & \text{if } i = j, \\ 0, & \text{if } i \in \{1, \dots, p\} \setminus \{j\}, \end{cases}$$

with $t > 0$, and $\lambda_i = 0$ for all $i \in \{1, \dots, m\}$. In this case, we obtain

$$t \to \infty \quad \Rightarrow \quad L(x, \lambda, \mu) = f(x) + th_j(x)^2 \to \infty.$$

(b) Assume that $g_i(x) > 0$ for some $i \in \{1, \dots, m\}$. Then, we can define

$$\lambda_j = \begin{cases} t, & \text{if } j = i, \\ 0, & \text{if } j \in \{1, \dots, m\} \setminus \{i\}, \end{cases}$$

with $t > 0$, and $\mu_j = 0$ for all $j \in \{1, \dots, p\}$. In this case, we have

$$t \to \infty \quad \Rightarrow \quad L(x, \lambda, \mu) = f(x) + tg_i(x) \to \infty.$$

From (a) and (b), we conclude that (2) also holds in this case.

□

The previous proposition shows that the original problem (NLP) can be written as

$$\begin{array}{ll} \min & \sup\limits_{(\lambda,\mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p} L(x, \lambda, \mu) \\ \text{s.t.} & x \in X, \end{array}$$

where

$$X = \left\{ x \in \mathbb{R}^n \;\middle|\; \sup_{(\lambda,\mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p} L(x, \lambda, \mu) < \infty \right\}.$$

By changing the order of the "min" and the "max" ("sup") in the above problem, we have:

$$\begin{array}{ll} \max & \omega(\lambda, \mu) \\ \text{s.t.} & (\lambda, \mu) \in D, \end{array} \tag{3}$$

where

$$D := \left\{ (\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p \mid \omega(\lambda, \mu) > -\infty \right\}.$$

and $\omega \colon D \to \mathbb{R}$ is defined by

$$\omega(\lambda, \mu) := \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu).$$

The maximization problem (3) is called the *dual problem* associated to (NLP).

Example 1. (Linear programming) *Let us show the dual of the following problem:*

$$\begin{array}{ll} \min & \langle c, x \rangle \\ \text{s.t.} & Ax \geq b, \end{array}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ *and* $b \in \mathbb{R}^m$ *are given.*

Proof. The above problem is a particular case of (NLP), with $f(x) = \langle c, x \rangle$, $g(x) = b - Ax$ and $h(x) = 0$. By ignoring the parts associated to the equality constraints, we have:

$$\begin{aligned} \omega(\lambda) &= \inf_{x \in \mathbb{R}^n} L(x, \lambda) \\ &= \inf_{x \in \mathbb{R}^n} \langle c, x \rangle + \langle b - Ax, \lambda \rangle \\ &= \inf_{x \in \mathbb{R}^n} \langle c - A^\top \lambda, x \rangle + \langle b, \lambda \rangle. \end{aligned}$$

We now show that $\omega(\lambda) > -\infty$ if and only if $c - A^\top \lambda = 0$. If $c - A^\top \lambda = 0$, we clearly obtain $\omega(\lambda) > -\infty$. Assume that $\omega(\lambda) > -\infty$ and $c - A^\top \lambda \neq 0$. If we define $x = -t(c - A^\top \lambda)$, then by taking $t \to \infty$, we have $\omega(\lambda) = -\infty$, which is a contradiction. Thus, the equivalence holds and the dual feasible set is written as

$$\begin{aligned} D &= \{\lambda \in \mathbb{R}^m_+ \mid \omega(\lambda) > -\infty\} \\ &= \{\lambda \in \mathbb{R}^m_+ \mid A^\top \lambda = c\}. \end{aligned}$$

Since $\omega(\lambda) = \langle b, \lambda \rangle$ for all $\lambda \in D$, we can write the dual problem as

$$\begin{array}{ll} \max & \langle b, \lambda \rangle \\ \text{s.t.} & A^\top \lambda = c, \\ & \lambda \geq 0. \end{array}$$

□

We now show that the dual problem has desirable properties, that does not depend on the problem's structure. In particular, (3) is a maximization of a concave objective function under a convex set. Therefore, all local optimal solutions of (3) are also global.

Proposition 2. *For any optimization problem of the type* (NLP) *and its dual problem* (3)*, the objective function* ω *is concave and the feasible dual set* D *is convex.*

Proof. First, let us note that the Lagrangian function L is linear with respect to the pair (λ, μ), that is,

$$\begin{aligned} & L(x, \alpha\lambda^1 + (1-\alpha)\lambda^2, \alpha\mu^1 + (1-\alpha)\mu^2) \\ = \; & \alpha L(x, \lambda^1, \mu^1) + (1-\alpha) L(x, \lambda^2, \mu^2) \end{aligned}$$

for all $\lambda^1, \lambda^2 \in \mathbb{R}^m$, $\mu^1, \mu^2 \in \mathbb{R}^p$ and $\alpha \in [0, 1]$. Therefore, we have

$$\begin{aligned} & \omega(\alpha\lambda^1 + (1-\alpha)\lambda^2, \alpha\mu^1 + (1-\alpha)\mu^2) \\ = \; & \inf_{x \in \mathbb{R}^n} L(x, \alpha\lambda^1 + (1-\alpha)\lambda^2, \alpha\mu^1 + (1-\alpha)\mu^2) \\ = \; & \inf_{x \in \mathbb{R}^n} \left[\alpha L(x, \lambda^1, \mu^1) + (1-\alpha) L(x, \lambda^2, \mu^2)\right] \\ \geq \; & \alpha \inf_{x \in \mathbb{R}^n} L(x, \lambda^1, \mu^1) + (1-\alpha) \inf_{x \in \mathbb{R}^n} L(x, \lambda^2, \mu^2) \\ = \; & \alpha\omega(\lambda^1, \mu^1) + (1-\alpha)\omega(\lambda^2, \mu^2), \end{aligned} \tag{4}$$

where the inequality holds because $\alpha \in [0,1]$ and by using a property of infimums. We then conclude that ω is concave. Now, let us define

$$\tilde{D} := \{(\lambda,\mu) \in \mathbb{R}^m \times \mathbb{R}^p \mid \omega(\lambda,\mu) > -\infty\}.$$

Observe that $D = \tilde{D} \cap (\mathbb{R}^m_+ \times \mathbb{R}^p)$. Letting $(\lambda^1,\mu^1),(\lambda^2,\mu^2) \in \tilde{D}$, we have $\omega(\lambda^1,\mu^1) > -\infty$ and $\omega(\lambda^2,\mu^2) > -\infty$. Then, taking $\alpha \in [0,1]$, we obtain $\alpha(\lambda^1,\mu^1) + (1-\alpha)(\lambda^2,\mu^2) \in \tilde{D}$ from (4), which means that $\tilde{D}$ is convex. Since $\mathbb{R}^m_+ \times \mathbb{R}^p$ is also convex, and D is an intersection of two convex sets, we conclude that D is also convex. □

The previous proposition shows that the dual problem is always a convex optimization problem (since maximization of a concave function under a convex set can be written as a minimization of a convex function under a convex set). However, the dual function ω is, in general, nondifferentiable, even if the primal problem is differentiable. Although the primal and dual problems are not equivalent in general, we can show that the dual problem can at least give an lower bound for the primal problem.

Theorem 1. (Weak duality) *For any optimization problem of the type* (NLP) *and its dual problem* (3)*, we have*

$$\omega(\lambda,\mu) \leq f(x)$$

for all feasible primal-dual pairs $x \in X$ and $(\lambda,\mu) \in D$. In particular, we obtain

$$\sup_{(\lambda,\mu)\in D} \omega(\lambda,\mu) \leq \inf_{x\in X} f(x). \tag{5}$$

Proof. Let $x \in X$ and $(\lambda,\mu) \in D$. Then,

$$\begin{aligned} \omega(\lambda,\mu) &= \inf_{z\in\mathbb{R}^n} L(z,\lambda,\mu) \\ &\leq L(x,\lambda,\mu) \\ &= f(x) + \langle g(x),\lambda\rangle + \langle h(x),\mu\rangle \\ &\leq f(x), \end{aligned}$$

where the second inequality holds because $h(x) = 0$, $g(x) \leq 0$ and $\lambda \geq 0$. The conclusion follows by taking the infimum and the supremum, respectively in the right-side and the left-side of the above inequality. □

When the inequality (5) is satisfied with an equality, we say that the *strong duality* holds. Otherwise, if the inequality is strict, then we say that there exists a *duality gap* between the problems. The strong duality holds only for particular optimization problems (e.g. linear programming). Still, for the general case, the weak duality theorem shows the following:

- If the primal problem is unbounded (i.e., $\inf_{x\in X} f(x) = -\infty$), then the dual problem is infeasible (i.e., $D = \emptyset$).
- If the dual problem is unbounded (i.e., $\sup_{(\lambda,\mu)\in D} \omega(\lambda,\mu) = \infty$), then the primal problem is infeasible (i.e., $X = \emptyset$).